A new digital image correlation algorithm for whole-field displacement measurement

C. SU and L. ANAND
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139, USA

Abstract— We have developed a new digital image correlation (DIC) algorithm for non-contact, two-dimensional, whole-field displacement and strain measurement. Relative to existing algorithms, our algorithm substantially reduces the calculation expense by using neighborhood information while processing the data to determine the displacement field in a sub-region of interest. The new algorithm also uses higher-order interpolations of the displacement field, allowing for better accuracy in estimating strain distributions when the deformation field is non-homogeneous. Numerically-generated digital images are used to show that the new algorithm accurately reproduces the imposed displacement fields. The algorithm is also tested on actual images from deformed specimens from a variety of experiments, and shown to perform satisfactorily.

Keywords— A. Experimental Mechanics B. Non-contact strain measurement. C. Mechanical testing

I. INTRODUCTION

The digital image correlation (DIC) method obtains the incremental displacement and strain field on the surface of a planar specimen by comparing a pair of digital images taken before and after the deformation. The idea of using digital images as a strain measurement technique, was first introduced by Peters et al. [1], and has since been widely used in experimental strain analysis [cf., e.g., 2-9].

As a strain measurement technique, the DIC method has many desirable properties: (a) It is non-contact, and is therefore suitable for measuring the deformation response of soft materials, such as polymeric and biomaterials. (b) It gives whole-field displacement and strain measurements, which is important when the deformation is non-homogeneous. (c) It is useful for strain measurement at the micro- and nanoscale from digital photomicrographs taken using techniques such as TEM, SEM, and AFM.

Commonly, images of a surface during deformation are recorded by digital photography. A digital image may be considered to be a matrix whose elements are positive integers which represent the surface brightness; each of these elements is called a pixel with a given value in a range which depends on the digital camera and the image acquisition electronics.\(^1\) By comparing successive images the 2D in-plane incremental displacement field is measured.\(^2\)

Consider an object with a flat surface, and let \(X\) be the position of a material point on the surface in the reference configuration at time \(t_0\), and \(x\) its position in the deformed configuration of the surface at a later time \(t > t_0\), then

\[
x = y(X, t) = X + u(X, t), \quad t > t_0, \quad (1)
\]

where \(y\) is the motion, and \(u\) is the displacement vector of the material point \(X\). During an experiment, images of the surface of the object are recorded at times \(t_0\) and \(t\). Let \(f(X, t_0)\) and \(g(x, t)\) represent the intensity distribution functions of the two images in the initial and deformed configurations, respectively. The principle of image correlation for deformation measurement is to find a displacement field \(u(X, t)\) that can match the two intensity distribution functions over a small but finite area \(S\) in the initial configuration, or the corresponding area \(S\) in the deformed configuration. The displacement field \(u\) over \(S\) is often parameterized by a vector \(P\)

\[
u(X, t) = u_P(X, P),
\]

\[
P = \{P_1(X_0, t), P_2(X_0, t), \ldots, P_N(X_0, t)\}, \quad (2)
\]

where \(X_0\) is a reference point within the subset \(S\). The match between the two images over the subset \(S\) for an assumed parameter set \(P\) is measured by a correlation coefficient

\[
C = C(P) = \tilde{C}(f(X, t_0), g(X + u_P, t)) = X_0, X \in S, \quad (3)
\]

with the correlation coefficient \(C\) and the displacement field \(u_P(X, P)\) defined over the subset \(S\) of the surface. The correlation coefficient \(C\) is often defined to have an extremum, typically a minimum, for the actual displacement field \(u(X, t)\) in the neighborhood of \(X_0\). Thus, the problem of determination of the displacement field by image correlation in the neighborhood of a point \(X_0\) located within the subset \(S\), may be stated as an optimization problem to find a set of local mapping parameters \(P^*\), such that

\[
C(P^*) = \min_{P \in P_D} C(P), \quad (4)
\]

where \(P_D\) is the domain of an assumed two-dimensional mapping function (cf., e.g. [7]).

A common correlation function is

\[
C = \frac{\sum_{S} (f-g)^2}{\sum_{S} f^2}, \quad (5)
\]
called the least-square correlation function. It is clear that it is non-negative and approaches 0 when \(f\) and \(g\) are close to each other.

\^1\For an 8-bit system it is [0, 255].
\^2\It is also possible to obtain full three-dimensional displacement fields using more sophisticated imaging techniques, such as AFM.
Now, expanding the deformation at time \( t \) about \( X_0 \), we obtain
\[
y(X, t) = y(X_0, t) + F(X_0, t)(X - X_0) + o(X - X_0),
\]
where \( F = \frac{\partial y}{\partial X} \) is the deformation gradient. Thus, in the neighborhood of a point \( X_0 \), and to within an error of \( o(X - X_0) \), a deformation behaves like a homogeneous deformation with deformation gradient \( F(X_0, t) \):
\[
y(X, t) = y(X_0, t) + F(X_0, t)(X - X_0).
\]
This may be equivalently written in terms of the displacement and displacement gradient as
\[
u(X, t) = u(X_0, t) + \frac{\partial u}{\partial X}(X_0, t)(X - X_0).
\]
Hence, once the displacement vector field \( \mathbf{u}(X, t) \) in the neighborhood of \( X_0 \) at time \( t \) has been calculated, all other quantities, such as the deformation gradient \( F(X_0, t) \), the stretch \( U(X_0, t) \), the rotation \( R(X_0, t) \), the finite Green strain \( E(X_0, t) \) tensors at \( (X_0, t) \) are easily calculated using standard definitions:
\[
F(X_0, t) = 1 + \frac{\partial u}{\partial X}(X_0, t),
\]
\[
U(X_0, t) = \sqrt{F(X_0, t)^T F(X_0, t)},
\]
\[
R(X_0, t) = F(X_0, t) U(X_0, t)^{-1},
\]
\[
E(X_0, t) = \frac{1}{2} [F(X_0, t)^T F(X_0, t) - 1].
\]

Writing \((X_0, Y_0)\) for the in-plane cartesian components of \( X_0 \), \((u, v)\) for the components of \( \mathbf{u}(X_0, t) \), and \((\frac{\partial u}{\partial X}, \frac{\partial u}{\partial Y}, \frac{\partial v}{\partial X}, \frac{\partial v}{\partial Y})\) for the components of \( \frac{\partial \mathbf{u}}{\partial X}(X_0, t) \), we may write (8) in expanded form as
\[
u(X, Y) = u_0 + \frac{\partial u}{\partial X}(X - X_0) + \frac{\partial u}{\partial Y}(Y - Y_0),
\]
\[
v(X, Y) = v_0 + \frac{\partial v}{\partial X}(X - X_0) + \frac{\partial v}{\partial Y}(Y - Y_0).
\]

In executing a DIC algorithm, the image of the reference configuration at time \( t_0 \) is divided into a mesh of subregions \( S \). Most current DIC algorithms (cf., e.g., [2,7]) assume that the deformation field in each subregion \( S \) with respect to its own reference point \( X_0 \) is homogeneous, and take the parameter set \( \mathbf{P} \) as
\[
\mathbf{P} = \left\{ u_0, v_0, \frac{\partial u}{\partial X}, \frac{\partial u}{\partial Y}, \frac{\partial v}{\partial X}, \frac{\partial v}{\partial Y} \right\}.
\]
The minimization problem (4) for each subregion is then typically solved using a coarse search to find a rough estimate for the solution vector \( \mathbf{P}^* \), and then the Newton-Raphson method is used to refine the solution. The \( n \)th partial correction for the Newton-Raphson method is given by
\[
\Delta \mathbf{P}_n = -\mathbf{H}^{-1}(\mathbf{P}_n) \times \nabla \mathbf{C}(\mathbf{P}_n)
\]
where in matrix notation,
\[
\Delta \mathbf{P} = \left[ dP_1, dP_2 \ldots dP_6 \right]^T,
\]
\[
\mathbf{H} = \left[ \frac{\partial^2 \mathbf{C}}{\partial P_i \partial P_j} \right]_{i, j = 1, 2 \ldots 6},
\]
\[
\nabla \mathbf{C}(\mathbf{P}) = \left( \frac{\partial \mathbf{C}}{\partial P_1}, \frac{\partial \mathbf{C}}{\partial P_2}, \ldots \frac{\partial \mathbf{C}}{\partial P_6} \right)^T.
\]

Calculation of the Hessian matrix \( \mathbf{H} \) is the most time-consuming step in the Newton-Raphson method. In the commonly used method described above, the vector \( \mathbf{P} \) has 6 components and the Hessian matrix \( \mathbf{H} \) has 36 components. In the next section we describe our new algorithm, which substantially reduces the calculation time for estimating the local displacement fields within the various subregions \( S \) in a mesh, using standard ideas of interpolations from finite element techniques.

II. NEW ALGORITHM

Our new algorithm also follows a coarse-fine search approach. The major difference is how it describes the local displacement field while carrying out the Newton-Raphson calculations. In the algorithm discussed in the previous section, the displacement field \( \mathbf{u}(X, t) \) in each subregion is determined locally; information from neighboring subregions is not utilized. However, when the deformation is continuous, the displacement field of each subregion has to satisfy compatibility requirements, and therefore the displacement fields of adjacent subregions are not totally independent.

Recall that in order to calculate the correlation function
\[
C = \sum_S \frac{(f(X, t_0) - g(X + \mathbf{u}(X, t), t))^2}{\sum_S f(X, t_0)^2},
\]
it is necessary to know the displacement field \( \mathbf{u}(X, t) \) at all points \( X \) in \( S \). In the new algorithm, this displacement field is determined by interpolating the displacement of the centroid \( X_0 \) of \( S \) and the centroid displacements of its eight surrounding subregions, see Fig. 1, as follows. Labelling the nine subregions as a \( 3 \times 3 \) matrix, the region \( S \) of interest is the subregion ‘22’. Let the coordinates of a point \( X \) in subregion ‘22’ at time \( t \) be denoted by \( (X = X_0^{22} + \Delta X, Y = Y_0^{22} + \Delta Y) \), where \( (X_0^{22}, Y_0^{22}) \) are the coordinates of its centroid \( X_0 \). Then, the components of the displacement field in the subregion of interest are interpolated as
\[
u^{(22)}(X, Y) = \sum_{i,j=1}^{3} h_{ij} u^{(ij)}_0,
\]
\[
v^{(22)}(X, Y) = \sum_{i,j=1}^{3} h_{ij} v^{(ij)}_0,
\]
where
\[
\left\{ u^{(ij)}_0, v^{(ij)}_0 \right\}, \quad i, j = 1, 2, 3,
\]
denote the components of the displacement of the centroids of the \( (ij) \)th subregion in a \( 3 \times 3 \) patch of nine subregions,
as depicted in Fig. 1, and $h_{ij}$ are interpolation functions given by

\begin{align}
h_{11} &= \frac{1}{4} r(1-r)s(1-s), \\
h_{12} &= -\frac{1}{2} r(1-r^2)s(1-s), \\
h_{13} &= -\frac{1}{4} r(1+r)s(1-s), \\
h_{21} &= -\frac{1}{2} r(1-r)(1-s^2), \\
h_{22} &= (1-r^2)(1-s^2), \\
h_{23} &= \frac{1}{2} r(1+r)(1-s^2), \\
h_{31} &= -\frac{1}{4} r(1-r)s(1+s), \\
h_{32} &= \frac{1}{2} (1-r^2)s(1+s), \\
h_{33} &= \frac{1}{4} r(1+r)s(1+s),
\end{align}

where the local coordinates $(r,s)$ are calculated as

\begin{align}
r &= \Delta X/L, \\
s &= \Delta Y/L,
\end{align}

with $L$ denoting the distance between the centers of two adjacent subregions.

We take the parameter set $\mathbf{P}$ for a subregion $S$ as the components of the displacement vector $(u_0, v_0)$ of its centroid $\mathbf{X}_0$, and continuing with the nomenclature above for the subregion ‘22’,

$$\mathbf{P} = \left\{ u_0 \equiv u_0^{(22)}, v_0 \equiv v_0^{(22)} \right\}.$$  \hfill (35)

With this parameter set, the $n^{th}$ partial corrections in the Newton-Raphson method are again given by

$$\Delta \mathbf{P}_n = -\mathbf{H}^{-1}(\mathbf{P}_n) \ast \nabla C(\mathbf{P}_n),$$  \hfill (36)

but now

\begin{align}
\Delta \mathbf{P} &= (du_0, dv_0)^T, \\
\mathbf{H} &= \begin{bmatrix}
\frac{\partial^2 C}{\partial u_0^2} & \frac{\partial^2 C}{\partial u_0 \partial v_0} \\
\frac{\partial^2 C}{\partial u_0 \partial v_0} & \frac{\partial^2 C}{\partial v_0^2}
\end{bmatrix}, \\
\nabla C(\mathbf{P}) &= \begin{bmatrix}
\frac{\partial C}{\partial u_0} & \frac{\partial C}{\partial v_0}
\end{bmatrix}^T.
\end{align}

\hfill (37) \hfill (38) \hfill (39)

Thus, we see that in the new algorithm the parameter set $\mathbf{P}$ and the Hessian matrix $\mathbf{H}$ have only 2 and 4 components, respectively, compared with 6 and 36 for the algorithm defined in the previous section. Further, we expect that the displacement field described by (21) and (22) will be a better representation of the displacement field described by (13) and (14) when the deformation is non-homogeneous, because the displacement fields of neighboring subregions are related to each other, and this information is being utilized in the new algorithm.

As an example of one of the derivatives for the Newton-Raphson procedure we note that one has to calculate derivatives of the type

$$\frac{\partial C}{\partial u_0} = -\frac{2}{\sum_s f(\mathbf{X}, t_0)^2} \sum_s \left[ (f(\mathbf{X}, t_0) - g(\mathbf{X} + \mathbf{u}(\mathbf{X}), t) \times \frac{\partial g(\mathbf{X} + \mathbf{u}(\mathbf{X}), t)}{\partial u_0} \right]$$ \hfill (40)
images which are subjected to two different deformation histories.

In order to use the DIC method, the surface of a specimen should either exhibit suitable natural random patterns, or it should be prepared by painting it with dots of random shapes, appropriate sizes and frequencies. It is possible to carry out such a surface preparation procedure digitally to create an artificial image to represent such a prepared surface. Further, such an image can also be manipulated digitally to impose any desired deformation field. The advantage of creating a sequence of deformed images digitally, is that the imposed deformation field can be of any form and precisely controlled, and any noise associated with imaging hardware and other sources of errors in physical experiments are eliminated.

The following two-dimensional polynomial function was used to generate a single two-dimensional ‘dot’:

\[
z = 1 - \frac{1}{c} \left[ a \times (x - x_0)^2 + b \times (y - y_0)^2 \right]^m. \tag{51}
\]

Here, the parameters \(a\) and \(b\) are random numbers between [0.9, 1.1] which give each dot a different ratio between its dimensions in the \(x\) and \(y\) directions; the parameter \(m\) controls the shape of the dot; and \(c\) controls the size of the dot for a specific \(m\). The reason for choosing random numbers for \(a\) and \(b\) is to make every dot distinct, so that later in the DIC calculation, the displacement solutions for the subregions will be reasonably different. This function has a value \(z\) in the range \((-\infty, 1]\). To convert a value of \(z\) for an image file, we need to multiply it by a suitable number, such as 255 for an 8-bit gray-scale image. Then, if the result is positive, its integer part is used in the image file, and if the result is negative, its value in the image file is set to zero. Fig. 3 illustrates the profile of such a dot and its image form. The values \(m = 2\) and \(c = 500\) were used to create this dot; this combination gives us a radius of roughly 5 pixels. A desired mesh of dots can be generated by successively repeating the procedure for a single dot.

A ‘deformation’ of the 2D polynomial function (51) can be achieved by transforming the original function. For example, the transformation

\[
z = 1 - \frac{1}{c} \left[ a \times (\lambda \cdot (x - \Delta x - x_0))^2 + b \times (y - y_0)^2 \right]^m, \tag{52}
\]

is equivalent to the surface having a rigid body motion \(\Delta x\) and a stretch \(\lambda\) in the \(x\) direction. Shear, rotation or any other kind of deformations can be imposed in such a fashion. Two different types of deformation fields have been used to test the accuracy of our DIC algorithm.

The first is a homogeneous simple shear. Fig. 4a shows the undeformed mesh of points representing a surface, and Fig. 4b shows the deformed geometry after a simple shear

\[
x = X + \gamma Y, \quad y = Y, \quad X, Y \in [0, 1], \tag{53}
\]

with \(\gamma = 0.2\). The deformation gradient tensor \(\mathbf{F}\) for this homogeneous deformation was calculated, and the polar decomposition of \(\mathbf{F}\) was performed to get the rotation \(\mathbf{R}\) and the stretch \(\mathbf{U}\), together with principal values \((\lambda_1, \lambda_2)\). Ten sequential frames were used to analyze the deformation. Fig. 4c shows a comparison of the principal stretches calculated using the DIC method against those calculated from the numerically imposed deformation (53), and Fig. 4d shows the corresponding calculations for the rotation, which for this two-dimensional deformation may be represented by an angle \(\theta\). In these figures, the full lines represent the theoretical value of the imposed deformation field, and the points are obtained from the new DIC algorithm. The results from the DIC calculation are in excellent agreement with those from the numerically-imposed deformation.

The simple non-homogeneous deformation field tested here is

\[
x = (1 + \epsilon) \left(X - \frac{1}{2}\right) + \frac{1}{2}, \quad y = Y, \quad X, Y \in [0, 1], \tag{54}
\]

\[
\epsilon = \epsilon(Y) = \epsilon_0 \left(4 \left(Y - \frac{1}{2}\right)^2 - 1\right), \quad \epsilon_0 = 0.1, \tag{55}
\]

with the following components for the two-dimensional infinitesimal strain measure,

\[
[\varepsilon] = \begin{bmatrix}
\epsilon & \frac{1}{2} \frac{\partial \epsilon}{\partial Y} X \\
\frac{1}{2} \frac{\partial \epsilon}{\partial X} X & 0
\end{bmatrix}. \tag{56}
\]

Fig. 5a shows the undeformed mesh of points representing a surface, and Fig. 5b shows the deformed geometry. Fig. 5c shows the variation of the numerically-imposed strain \(\varepsilon_{XX} = \varepsilon(Y)\) and compares it to the data calculated using the new DIC algorithm; as before, the solid line represents the theoretical value and the points are obtained from the DIC algorithm. We see that the DIC result matches the imposed strain variation very well.

Finally, Figs. 6a,b show the initial and deformed geometries for a mesh subjected to a similar non-homogeneous deformation, but this time for only one-quarter of the specimen. Figs. 6c,d show the corresponding distributions of \(\varepsilon_{XX}\) and \(\varepsilon_{XY}\) calculated using the DIC method. These distributions also match the theoretically-imposed ones, but the latter are not plotted in Figs. 6c,d for clarity.

IV. Experimental results

First we demonstrate the capability of our DIC method in a large strain application of determining the axial stretch during simple extension of a rubbery polymer. Because of the typical inhomogeneity of the deformation that develops in such experiments near the grips, the cross-head extension cannot be used to calculate the actual strain in the gauge section. Further, since the material is relatively soft and subjected to large strains, mechanical extensometers are also not suitable for strain measurements in such experiments. However, the non-contact DIC method may be conveniently used in this application.

Fig. 7a shows the initial surface of a tensile specimen of this material; the gauge length for calculating the axial stretch is also indicated in this figure. This time the surface
of the specimen is physically painted with a mesh of dots. The tensile test is carried out under displacement control. An image was taken every second, while the axial load $F$ was also recorded at the same frequency. Fig. 7b shows the specimen after a maximum axial stretch of $\lambda = 2.25$, which corresponds to an engineering strain of 125%. After processing the images with the new DIC method we obtained the stretch versus time curve $\lambda(t)$; combining it with the load versus time curve $F(t)$, we obtained the load-stretch curve $F(\lambda)$ shown in Fig. 7c. From the load-stretch data, and knowledge about the specimen geometry, and the standard assumption of incompressibility for rubber-like materials, other quantities such as true stress and true strain are easily calculated.

As a second example, we present results from an effort to determine the incremental strain distribution beneath a sharp indenter impressed on another rubbery polymer. The indentation was conducted in a scanning electron microscope (SEM) and the high-magnification images before and after the deformation were obtained by using this instrument. Fig. 8a,b. By comparing these two images using the new DIC method, the strain field in the polymeric material under the indenter was calculated. In this test, the surface of the specimen was not specially treated; the natural features on the surface of the polymeric material provided enough information for us to use the DIC method. Figs. 8c,d, respectively, show the calculated contours of the strain components $E_{XX}$ and $E_{YY}$ of the finite Green strain measure $E$. These figures clearly illustrate the concentration of the strain field underneath the indenter. The scale bar in Fig. 8a shows that the total region of interest here is only a few hundred square microns; it is difficult to measure the strain distribution in such a small region using any other current experimental method. However, the DIC method, combined with the SEM digital imaging capability, is readily useful for such an application at the microscale.

V. Concluding remarks

We have developed a new, reasonably fast, DIC algorithm for non-contact, two-dimensional, whole-field displacement and strain measurement. We have used numerically-generated digital images to show that the new algorithm accurately reproduces imposed homogeneous and non-homogeneous strain fields. The algorithm has also been used to measure the strain distributions from actual images obtained from deformed specimens, one at the macroscale of the stretch measurement of a tension test in an elastomeric material, and another at the microscale for the sharp indentation of another elastomeric material carried out in a SEM. In these two different challenging applications, the new DIC algorithm demonstrates some of its versatile attributes, such as a non-contact capability to measure large strains at the macroscale, and whole-field strain measurements at the microscale.

We have not carried out an exhaustive comparative study of the speed and accuracy of our algorithms versus existing algorithms for whole-field strain measurements in the literature. We leave such a study for future research.

Acknowledgments

The authors acknowledge the financial support provided by the Singapore-MIT Alliance, as well as ONR Contract N00014-01-1-0808 with MIT. The gracious help of Professor D. P.Hart of MIT in getting us started with digital image correlation methods is gratefully acknowledged.

References

Fig. 3. Artificially created image: (a) profile of a single dot; (b) image of a single dot.

Fig. 4. Homogeneous simple shear deformation: (a) initial image; (b) deformed image; (c) variation of principal stretches $\lambda_1, \lambda_2$, with increasing amount of shear $\gamma$; (d) variation of angle $\theta$ describing the rotation $R$ in simple shear, with increasing amount of shear.
Fig. 5. Non-homogeneous deformation: (a) initial image; (b) deformed image; (c) distribution of strain $\epsilon_{XX}$ versus $Y$ in the non-homogeneous deformation field.

Fig. 6. Non-homogeneous deformation: (a) initial image; (b) deformed image; (c) plot of distribution of strain $\epsilon_{XX}$ versus $X$ and $Y$; (d) plot of distribution of strain $\epsilon_{XY}$ versus $X$ and $Y$. 
Fig. 7. Tensile test of a rubber specimen to a large strain: (a) initial image showing the gauge length used to measure the axial stretch; (b) final image; (c) calculated force versus stretch curve.

Fig. 8. Strain distribution in microscale sharp indentation of a rubbery polymer: (a) initial image showing the white-framed area in which the incremental strain distribution is to be determined; (b) final deformed image; the images are obtained in a SEM. (c,d) Calculated contours of the $E_{XX}$ and $E_{YY}$ components, respectively, of the finite Green strain measure.